Automatic String Detection for Bass Guitar and Electric Guitar

Jakob Abeßer

Fraunhofer IDMT,
Ehrenbergstr. 17, 98693 Ilmenau, Germany
abr@idmt.fraunhofer.de
http://www.idmt.fraunhofer.de

Abstract. In this paper, we present a machine learning-based approach to automatically estimate the fretboard position (string number and fret number) from recordings of the bass guitar and the electric guitar. We perform different experiments to evaluate the classification performance on isolated note recordings. First, we analyze how the separation of training and test data in terms of instrument, playing-style, and pick-up setting affects the algorithm's performance. Second, we investigate how the performance can be improved by rejecting implausible classification results and by aggregating the classification results over multiple time frames. The algorithm showed highest string classification f-measure values of $F = .93$ for the bass guitar (4 classes) and $F = .90$ for the electric guitar (6 classes). A listening test with 9 participants with classification scores of $F = .26$ and $F = .16$ for bass guitar and electric guitar confirmed that the given tasks are very challenging to human listeners. Finally, we discuss further research directions with special focus on the application of automatic string detection in music education and software.

Keywords: string classification, fretboard position, fingering, bass guitar, electric guitar, inharmonicity coefficient.

1 Introduction

On string instruments such as the bass guitar or the guitar, most notes within the instrument's pitch range can be played at multiple positions on the instrument fretboard. The *fretboard position* of a note is defined by the string number n_s and the fret number n_f. Common music notation such as the *score* does not provide any information about the fretboard positions to be applied. Instead, musicians often have to choose appropriate fretboard positions based on their musical experience and stylistic preferences. The *tablature* notation, on the other hand, is specialized for the geometry of fretted string instruments such as the guitar or the bass guitar. It specifies the string and fret number for each note and thus resolves the ambiguity between note pitch and fretboard position. Figure 1 shows a bass-line both as score and tablature notation.

M. Aramaki et al. (Eds.): CMMR 2012, LNCS 7900, pp. 333–352, 2013.

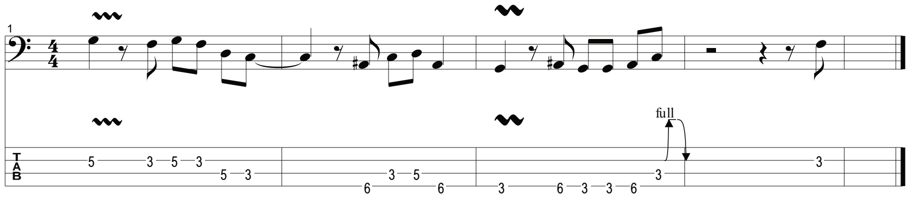


Fig. 1. Score and tablature notation of a bass-line. The four horizontal lines in the tablature notation correspond to the four strings with the tuning E1, A2, D2, and G2 (from bottom to top). The numbers correspond to the fret numbers on the strings that are to be played.

Conventional automatic music transcription algorithms only extract *score-related note parameters* such as pitch, onset, and duration. In order to analyze recordings of string instruments, the string and fret number need to be estimated as additional *instrument-related note parameters.* Algorithms for automatic tablature generation from an audio recordings can be applied in music assistance and music education software. Tablature notations are especially helpful to novices who are not familiar with reading musical scores.

As will be discussed in Section 3, various methods for estimating the fretboard position were proposed in the literature so far, ranging from pure audio-based methods to methods that exploit the visual modality or methods that use attached sensors on the instrument. However, the exclusive focus on audio analysis methods for this purpose bears several advantages: in music performance scenarios involving a bass guitar or an electric guitar, the instrument signals are directly accessible from the instrument's output jack. In contrast, video recordings of performing musicians and the instrument neck are often limited in quality due to movement, shading, and varying lighting conditions on stage. Additional sensors or cameras that need to be attached to the instrument are often obtrusive to the musicians and affect their musical performance. Therefore, we focus on audio-based analysis in this paper.

This paper is structured as follows: In Section 2, we outline the goals and challenges of this work. In Section 3, we discuss existing methods for estimating the fretboard position from string instrument recordings. We introduce a novel audio-based approach in Section 4, starting with the spectral modeling of recorded bass and guitar notes in Section 4.1. Based on the audio features explained in Section 4.2, we illustrate how the fretboard position is automatically estimated in Section 4.3. In Section 5, we present several experiments to evaluate the algorithm's performance and discuss the obtained results. This section also includes the results of a listening test with human participants for the task of string classification. Finally, we conclude our work in Section 6 and give an outlook in Section 7 on future research.

2 Goals and Challenges

In this paper, we aim to estimate the string number n_s from notes recorded with bass guitars and electric guitars. Based on the note (MIDI) pitch P and the string number, we can apply knowledge on the instrument tuning to derive the fret number n_f. In the evaluation experiments described in Section 5, we investigate how the classification results are affected by separating the training and test data according to different criteria such as the instruments, the pick-up (PU) settings, and the applied playing techniques. Furthermore, we analyze if a majority voting scheme that combines multiple string classification results for each note can improve the classification performance. Finally, the obtained results are compared to the human performance for the same task.

The main challenge of this work is to identify suitable audio features that allow to discriminate between notes that, on the one hand, have the same fundamental frequency f_0 but, on the other hand, are played on different strings. The automatic classification of the played string is difficult since the change of fingering alters the sonic properties of the recorded music signal only subtly. This was confirmed in the human listening test presented in Section 5.2.

Classic non-parametric spectral estimation techniques such as the Short-time Fourier Transform (STFT) are affected by the spectral leakage effect: the Fourier Transform of the applied window function limits the achievable frequency resolution to resolve the exact frequency position of spectral peaks. In order to achieve a sufficiently high frequency resolution for estimating the harmonic frequencies of a note, rather larger time frames are necessary. The decreased time resolution is disadvantageous if notes are played with frequency modulation techniques such as bending or vibrato, which cause short-term fluctuations of the harmonic frequencies [1]. This problem is especially impeding in lower frequency bands.

Thus, a system based on non-parametric spectral estimation techniques is only applicable to analyze notes with no or only slow pitch variation. This can be a severe limitation for a real-world application scenario such as music education software. Since we focus on the bass guitar and the electric guitar, frequencies between 41.2 Hz and 659.3 Hz need to be investigated as potential f_0-candidates[1].

3 Related Work

In this section, we discuss previous work on the estimation of the played string and the fretboard position from bass and guitar recordings. First, we review methods that solely focus on analyzing the audio signal. Special focus is given to the analysis of inharmonicity. Then, we compare different hybrid methods that incorporate computer vision techniques, instrument enhancements, and sensors.

[1] This corresponds to the most commonly used bass guitar string tunings E2 to G3 and electric guitar string tuning E3 to E5, respectively, and a fret range up to the 12th fret position.

3.1 Audio Analysis

Penttinen et al. estimated the plucking point on a string by analyzing the delay times of the two waves on the string, which travel in opposite directions after the string is plucked [22]. This approach solely focuses on a time-domain analysis and is limited to monophonic signals.

In [3], Barbancho et al. presented an algorithm to estimate the string number from isolated guitar note recordings. The instrument samples used for evaluation were recorded using different playing techniques, different dynamic levels, and guitars with different string material. After the signal envelope is detected in the time-domain, spectral analysis based on STFT is applied to extract the spectral peaks. Then, various audio features related to the timbre of the notes are extracted such as the spectral centroid, the relative harmonic amplitudes of the first four harmonics, and the inharmonicity coefficient (see also Section 3.1). Furthermore, the temporal evolution of the partial amplitudes is captured by fitting an exponentially decaying envelope function. Consequently, only one feature vector can be extracted for each note. As will be shown in Section 4.2, the presented approach in this paper allows us to extract a single feature vectors for each time frame. This allows us to accumulate classification results from multiple feature vectors that were obtained from the same note recording to improve the classification performance (compare Section 4.3). The authors of [3] reported diverse results from the classification experiments. However, they did not provide an overall performance measure to compare against. The performance of the applied classification algorithm strongly varied for different note pitch values as well as for different compilations of the training set in their experiments.

In [2], Barbancho et al. presented a system for polyphonic transcription of guitar chords, which also allows estimation of the fingering of the chord on the guitar. The authors investigated 330 different fingering configurations for the most common three-voiced and four-voiced guitar chords. A Hidden Markov Model (HMM) is used to model all fingering configurations as individual hidden states. Based on an existing multi-pitch estimation algorithm, harmonic saliency values are computed for all possible pitch values within the pitch range of the guitar. Then, these saliency values are used as observations for the HMM. The transitions between different hidden states are furthermore constrained by two models—a musicological model, which captures the likelihood of different chord changes, and an acoustic model, which measures the physical difficulty of changing the chord fingerings. The authors emphasized that the presented algorithm is limited to the analysis of solo guitar recordings. However, in that scenario, the algorithm clearly outperformed a state-of-the-art chord transcription system. The applied dataset contained instrument samples of electric guitar and acoustic guitar.

Maezawa et al. proposed a system for automatic string detection from isolated bowed violin note recordings in [17]. Similar to the bass guitar, the violin has 4 different strings, but within a higher pitch range. The authors analyzed monophonic violin recordings of various classical pieces with given score information. First, the audio signal is temporally aligned to the musical score. For the string classification, filterbank energies are used as audio features and a Gaussian mixture model (GMM) is applied as classifier. The authors proposed two additional steps to increase the robustness of the classification. First, feature averaging and feature normalization are used. Then, a context-dependent error correction is applied, which is based on empirically observed rules describing how musicians choose the string number. The authors investigated how training and testing with the same and different instruments and string types affect the classification scores (similar to Section 5). The highest F-measure value that was achieved for the string classification with 4 classes is $F = .86$.

In [4], Barbancho et al. presented an algorithm for automatic tablature generation from audio recordings of guitar. First, one or multiple fundamental frequencies are detected by investigating the most prominent peaks as f_0-candidates. Each candidate is rated based on the fitness of the corresponding partial peaks to a given model that incorporates inharmonicity. The string and fret number of the detected notes are taken from the best fitting model parameters. Multiple notes are obtained by iteratively removing detected fundamental frequency and harmonic components from the spectrum. For the analysis of guitar chords, the authors focus on two scenarios: guitar chords of arbitrary shape with up to 4 chord notes, and guitar chords with a known (template) shape such as barré chords with up to 6 chord notes. The authors also use constraints to avoid note combinations that exceed the hand span of a musician and thus cannot be played on the guitar neck. The presented algorithm performed well for the fretboard detection of single notes with error rates between 0 and 0.11 for instrument samples of the RWC database and samples recorded by the authors themselves.

Inharmonicity. For musical instruments such as the piano, the guitar, or the bass guitar, the equation describing the vibration of an ideal flexible string is extended by a restoring force caused by the string stiffness [8]. Due to dispersive wave propagation within the vibrating string, the effect of inharmonicity occurs, i.e., the purely harmonic frequency relationship of an ideal string is distorted and the harmonic frequencies are stretched towards higher values as

$$f_k = k f_0 \sqrt{1 + \beta k^2};\ k \geq 1 \tag{1}$$

with k being the harmonic index of each overtone and f_0 being the fundamental frequency. The inharmonicity coefficient β depends on different properties of the vibrating string such as Young's Modulus E, the radius of gyration K, the string tension T, the cross-sectional area S, as well as the string length L. With the string length being approximately constant for all strings of the bass guitar and the electric guitar, the string diameter usually varies from 0.45 mm to 1.05 mm for electric bass and from 0.1 mm to 0.41 mm for electric guitar[2]. The string tension T is proportional to the square of the fundamental frequency of the vibrating string. Järveläinen et al. performed different listening tests to investigate the audibility of inharmonicity towards humans [13]. They found

[2] These values correspond to commonly used string gauges.

that the human audibility threshold for inharmonicity increases with increasing fundamental frequency.

Hodgekinson et al. observed a systematic time-dependence of the inharmonicity coefficient if the string is plucked hard [11]. The authors found that β does not remain constant but increases over time for an acoustic guitar note. In contrast, for a piano note, no such behavior was observed. In this paper, we aim to estimate β within single spectral frames and therefore do not take the temporal evolution of β into account.

Different methods have been applied in the literature to extract the inharmonicity coefficient such as the cepstral analysis, the harmonic product spectrum [9], or inharmonic comb-filter [10]. For the purpose of sound synthesis, especially for physical modeling of string instruments, inharmonicity is often included into the synthesis models in order to achieve a more natural sound [25].

3.2 Hybrid Approaches and Visual Approaches

Different methods for estimating the fretboard position from guitar recordings have been presented in the literature including analysis methods from computer vision as a multi-modal extension of audio-based analysis.

A combined audio and video analysis was proposed by Hybryk and Kim to estimate the fretboard position of chords that were played on an acoustic guitar [12]. The goal of this paper was to first identify a played chord on the guitar in terms of its chord style, i.e., root note and musical mode such as minor or major. For this purpose, the Specmurt [23] algorithm was used for spectral analysis in order to estimate a set of fundamental frequency candidates that can be associated with different note pitches. Based on the computed chord style (e.g., E minor), the chord voicing was estimated by tracking the spatial position of the hand on the instrument neck. The chord voicing is similar to the chord fingering as described in [2].

Another multi-modal approach for transcribing acoustic guitar performances was presented by Paleari et al. in [20]. In addition to audio analysis, the visual modality was analyzed to track the hand of the guitar players during their performance to estimate the fretboard position. The performing musicians were recorded using both two microphones and a digital video camera. The fretboard was first detected and then spatially tracked over time.

Other approaches solely used computer vision techniques for spatial transcription. Burns and Wanderley presented an algorithm for real-time finger-tracking in [5]. They used *attached cameras* on the guitar in order to get video recordings of the playing hand on the instrument neck. Kerdvibulvech and Saito used a stereo-camera setup to record a guitar player in [14]. Their system for finger-tracking requires the musician to wear *colored fingertips*. The main disadvantage of all these approaches is that both the attached cameras as well as the colored fingertips are unnatural for the guitar player. Therefore, they likely limit and impede the musician's expressive gestures and playing style.

Enhanced music instruments are equipped with additional sensors and controllers in order to directly measure the desired parameters instead of estimating

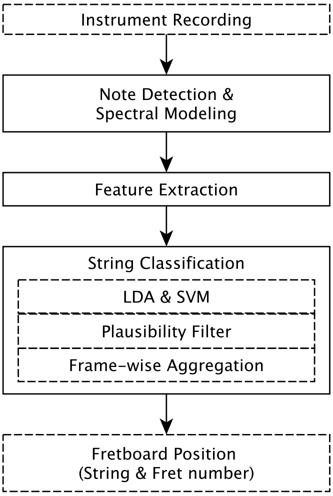


Fig. 2. Algorithm overview

them from the audio or video signal. On the one hand, these approaches lead to a high detection accuracy. On the other hand, these instrument extensions are obstructive to the musicians and can affect their performance on the instrument [12]. In contrast to regular electric guitar pickups, *hexaphonic pickups* separately capture each vibrating string. In this way, spectral overlap between the string signals is avoided, which allows a fast and robust pitch detection with very low latency and very high accuracy, as shown for instance by O'Grady and Rickard in [19].

4 Proposed System

Figure 2 provides an overview over the string classification algorithm proposed in this paper. All processing steps are detailed in the next sections.

4.1 Spectral Modeling

Non-parametric spectral estimation methods such as the Periodogram make no explicit assumption on the type of signal that is analyzed. In order to obtain a high frequency resolution for precise f_0-detection, relatively large time frames of data samples are necessary in order to compensate the spectral leakage effect, which is introduced by windowing the signal into frames. In contrast to the percussive nature of its short attack part (between approx. 20 ms and 40 ms), the decay part of a plucked string note can be modeled by a sum of decaying sinusoidal components. Their frequencies have a nearly perfectly harmonic relationship. Since the strings of the bass guitar and the electric guitar have a certain amount of stiffness, the known phenomenon of inharmonicity appears (compare Section 3.1).

Parametric spectral estimation techniques can be applied if the analyzed signal can be assumed to be generated by a known model. In our case, the power spectral density (PSD) $\Phi(\omega)$ can be modeled by an auto-regressive (AR) filter such as

$$\Phi(\omega) \approx \Phi_{AR}(\omega) = \sigma^2 \left| \frac{1}{1 + \sum_{l=1}^{p} a_l e^{-jl\omega}} \right|^2 \qquad (2)$$

with σ^2 denoting the process variance, p denoting the model order, and $\{a_l\} \in \mathbb{R}^{p+1}$ being the filter coefficients. Since auto-regressive processes are closely related to linear prediction (LP), both a *forward prediction error* and a *backward prediction error* can be defined to measure the predictive quality of the AR filter. We use the *least-squares method* (also known as *modified covariance method*) for spectral estimation. It is based on a simultaneous least-squares minimization of both prediction errors with respect to all filter coefficients $\{a_l\}$. This method has been shown to outperform related algorithms such as the Yule-Walker method, the Burg algorithm, and the covariance method (See [18] for more details). The size of the time frames N is only restricted by the model order as $p \leq 2N/3$.

First, we down-sample the signals to $f_s = 5.5$ kHz for the bass guitar samples and $f_s = 10.1$ kHz for the electric guitar samples. This way, we can detect the first 15 harmonics of each note within the instrument pitch ranges, which is necessary for the subsequent feature extraction as explained in Section 4.2. In Figure 3, the estimated AR power spectral density for a bass guitar sample (E1) as well as the estimated partials are illustrated. Within this paper, we compute the fundamental frequency f_0 from the known fretboard position of all notes in the dataset. The separate evaluation of fundamental frequency estimation is not within the scope of this paper.

By using overlapping time frames with a block-size of $N = 256$ and a hop-size of $H = 64$, we apply the spectral estimation algorithm to compute frame-wise estimates of the filter coefficients $\{a_l(n)\}$ in the frames that are selected for analysis (compare Section 4.2). In order to estimate the harmonic frequencies $\{f_k\}$, we first compute the pole frequencies of the AR filter by computing the roots of the numerator in Equation (2). Then, we assign one pole frequency to each harmonic according to the highest proximity to its theoretical frequency value as computed using Equation (1).

4.2 Feature Extraction

Note Detection. In Section 4.1, we discussed that notes played on the bass guitar and the guitar follow a signal model of decaying sinusoidal components, i.e., the partials. In this section, we discuss how we extract audio features that capture the amplitude and frequency characteristics. We first detect the first frame shortly after the note attack part of the note is finished and the harmonic decay part begins. As mentioned in Section 4.1, signal frames with a percussive

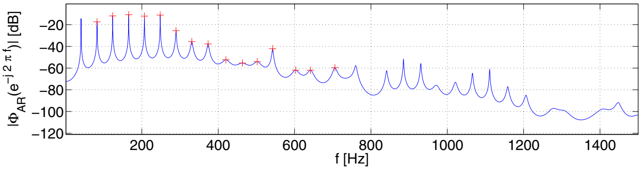


Fig. 3. Estimated AR power spectral density for the bass guitar sample with pitch E1 ($f_0 = 44.1Hz$). The estimated first 15 partials are indicated with red crosses.

characteristic are indicated by high values of the process variance $\sigma^2(n)$ obtained the AR spectral estimation. We found that time frames after

$$n^\star = \arg\max_n \sigma^2(n) \tag{3}$$

are suitable for feature extraction. If the aggregation of multiple frame-wise results is used, we extract features in the first 5 frames after $n^\star$. If the aggregation is not applied, one feature vector is computed for each note in the first frame after $n^\star$.

Inharmonicity Estimation. In each analyzed frame, we estimate the discrete frequencies f_k of the first 15 partials. Then, we estimate the inharmonicity coefficient β_k as follows. From Equation (1), we obtain

$$(f_k/f_0)^2 = k^2 + \beta k^4 \tag{4}$$

We use polynomial curve fitting to approximate the left-hand side of Equation (4) by a polynomial function of order 4 as

$$(f_k/f_0)^2 \approx \sum_{i=0}^{4} p_i k^i \tag{5}$$

and use the coefficient p_4 as an estimate of the inharmonicity coefficient β:

$$\hat{\beta} \approx p_4 \tag{6}$$

Partial-Based Features. In addition to the inharmonicity coefficient β, we compute various audio features that capture the amplitude and frequency characteristics of the first 15 partials of a note. First, we compute the relative amplitudes

$$\{\hat{a}_{r,k}\} = \{a_k/a_0\} \tag{7}$$

of the first 15 partials related to the amplitude of the fundamental frequency. Then, we approximate the relative partial amplitude values $\{\hat{a}_{r,k}\}$ as a linear function over k as

$$\hat{a}_{r,k} \approx p_1 k + p_0 \tag{8}$$

Table 1. Overview of all applied audio features

Feature	Number of dimensions
Inharmonicity coefficient $\hat{\beta}$	1
Relative partial amplitudes $\{\hat{a}_{r,k}\}$	15
Statistics over $\{\hat{a}_{r,k}\}$	8
Normalized partial frequency deviations $\{\Delta\hat{f}_{norm,k}\}$	15
Statistics over $\{\Delta\hat{f}_{norm,k}\}$	8
Partial amplitude slope $\hat{s}_a$	1
All features	$\sum = 48$

by using linear regression. We use the feature $\hat{s}_a = p_1$ as estimate of the *spectral slope* towards higher partial frequencies.

Based on the estimated inharmonicity coefficient $\hat{\beta}$ and the fundamental frequency f_0, we compute the theoretical partial frequency values $\{f_{k,theo}\}$ of the first 15 partials based on Equation (1) as

$$f_{k,theo} = kf_0\sqrt{1 + \hat{\beta}k^2}. \tag{9}$$

Then, we compute the deviation between the theoretical and estimated partial frequency values and normalize this difference value as

$$\Delta\hat{f}_{norm,k} = \frac{f_{k,theo} - \hat{f}_k}{\hat{f}_k}. \tag{10}$$

Again, we compute $\{\Delta\hat{f}_{norm,k}\}$ for the first 15 partials and use them as features. In addition, we compute the statistical descriptors: maximum value, minimum value, mean, median, mode (most frequent sample), variance, skewness, and kurtosis over both $\{\hat{a}_{r,k}\}$ and $\{\Delta\hat{f}_{norm,k}\}$. Table 1 provides an overview over all features and their dimensionality.

4.3 Estimation Of the Fretboard Position

String Classification. In order to automatically estimate the fretboard position from a note recording, we first aim to estimate the string number n_s. Therefore, we compute the 48-dimensional feature vector $\{x_i\}$ as described in the previous section. We use Linear Discriminant Analysis (LDA) to reduce the dimensionality of the feature space to $N_d = 3$ dimensions for bass guitar (4 string classes) and to $N_d = 5$ dimensions for guitar (6 string classes)[3], respectively. Then we train a Support Vector Machine (SVM) classifier using a Radial Basis Function (RBF) kernel with the classes defined by notes played on each string. SVM is a binary discriminative classifier that attempts to find an optimal

[3] The number of dimensions N_d is chosen as $N_d = N_{strings} - 1$.

decision plane between feature vectors of the different training classes [26]. The two kernel parameters C and γ are optimized based on a three-fold grid search. We use the LIBSVM library for our experiments [6].

The SVM returns the posterior probability $\{p\}$ for each string class. We estimate the string number $\hat{n}_s$ by maximizing p_i as

$$\hat{n}_s = \arg\max_i p_i. \tag{11}$$

We derive the fret number $\hat{n}_f$ from the estimated string number $\hat{n}_s$ by using knowledge of the instrument tuning as follows. The common tuning of the bass is E1, A2, D2, and G2; the tuning of the guitar is E2, A2, D3, G3, B3, and E4. The string tunings can be directly translated into a vector of corresponding MIDI pitch values as $\{P_S\} = [28, 33, 38, 43]$ and $\{P_S\} = [40, 45, 50, 55, 59, 64]$, respectively.

In order to derive the fret number $\hat{n}_s$, we first obtain the MIDI pitch value P that corresponds to the fundamental frequency $\hat{f}_0$ as

$$\hat{P} = \lfloor 12 \log_2(\hat{f}_0/440) - 69 \rfloor \tag{12}$$

Given the estimated string number $\hat{n}_s$, the fret number can be computed as

$$\hat{n}_f = \hat{P} - P_S(\hat{n}_s). \tag{13}$$

A fret number of $\hat{n}_f = 0$ indicates that a note was played by plucking an open string.

Plausibility Filter. As mentioned earlier, most note pitches within the frequency range of both the bass guitar and the guitar can be played on either one, two, or three different fret positions on the instrument neck. The total instrument pitch range is E2 to G3 for the bass guitar and E3 to E5 for the electric guitar[4]. Based on knowledge about the instrument string tunings, we can derive a set of MIDI pitch values that can be played on each string. Therefore, for each estimated MIDI pitch value $\hat{P}$, we can derive a list of strings on which this note can theoretically be played. If the plausibility filter is used, before estimating the string number as shown in Equation (11), we set the probability values in $\{p_i\}$ to zero for all strings, on which this note can not be played on.

Aggregation of Multiple Classification Results. If the result aggregation is used, all class probability values $\{p\}$ are summed up over 5 adjacent time frames and then again normalized to unit sum. The string number is then estimated by applying Equation (11) on the accumulated probability values.

[4] Here, a limited fret range up to the 12th fret position is considered in the database.

5 Evaluation and Results

5.1 Dataset

For the evaluation experiments, a dataset of 1034 audio samples was used. These samples are isolated note recordings, which were taken from the dataset previously published in [24].[5] The samples were recorded using two different bass guitars and two different electric guitars, each played with two different plucking styles (plucked with a plectrum and plucked with the fingers) and recorded with two different pick-up settings (either neck pick-up or body pick-up).

5.2 Experiments and Results

Experiment 1: Feature Selection for String Classification. In the first experiment, we aim to identify the most discriminatory features for the automatic string classification task as discussed in Section 4.3. For this purpose, the feature selection algorithm Inertia Ratio Maximization using Feature Space Projection (IRMFSP) [16, 21] was applied to all feature vectors and their corresponding class labels. This experiment was performed separately for both instruments. Table 2 lists the five most discriminatory features that were first selected by the IRMFSP algorithm for the bass guitar and the electric guitar.

The features $\Delta\hat{f}_{norm}$, $\hat{\beta}$, and $\hat{a}_{r,k}$ as well as the derived statistic measures were selected consistently for both instruments. These features measure frequency and amplitude characteristics of the partials and show high discriminative power between notes played on different strings independently of the applied instrument. The boxplots of the two most discriminative features $\Delta f_{norm,9}$ for bass and $\Delta f_{norm,15}$ for guitar are illustrated separately for each instrument string in Figure 4.

Since the deviation of the estimated harmonic frequencies from their theoretical values seems to carry distinctive information to discern between notes on different instrument strings, future work should investigate whether Equation (1) could be extended by higher order polynomial terms in order to better fit to the estimated harmonic frequency values.

Experiment 2: String Classification in Different Conditions. In this experiment, we aim to investigate the influence of

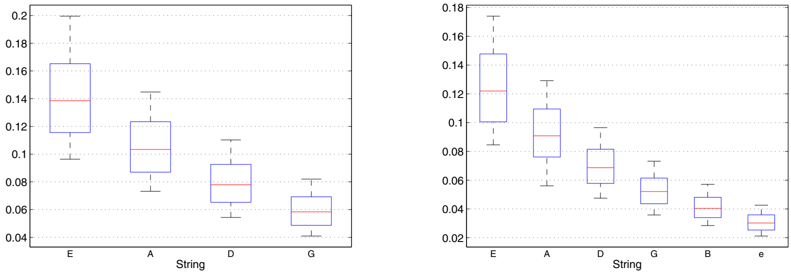
- the separation of the training and test set according to the applied instrument, playing technique, and pick-up setting,
- the instrument / the number of string classes,
- the use of a plausibility filter (compare Section 4.3),
- and the use of a aggregation of multiple classification results for each sample (compare Section 4.3).

on the performance of the automatic string classification algorithm.

[5] This dataset contains isolated notes from bass guitar and electric guitar processed with various audio effects. In this work, only the non-processed note recordings were used.

Table 2. Most discriminative audio features for the string classification task as discussed in Section 5.2. Features are given in order as selected by the IRMFSP algorithm.

Rank	Bass Guitar	Electric Guitar
1	$\Delta\hat{f}_{norm,9}$	$\Delta\hat{f}_{norm,15}$
2	$\hat{\beta}$	$\mathrm{mean}\{\hat{a}_{r,k}\}$
3	$\Delta\hat{f}_{norm,3}$	$\mathrm{var}\{\Delta\hat{f}_{norm,k}\}$
4	$\mathrm{var}\{\Delta\hat{f}_{norm,k}\}$	$\max\{\hat{a}_{r,k}\}$
5	$\hat{a}_{r,4}$	$\mathrm{skew}\{\Delta\hat{f}_{norm,k}\}$



(a) Boxplot of feature $\Delta f_{norm,9}$ for bass.

(b) Boxplot of feature $\Delta f_{norm,15}$ for guitar.

Fig. 4. Boxplots of the two most discriminative features for bass guitar and electric guitar

The different experiment conditions are illustrated in Table 3 for the bass guitar and in Table 4 for the electric guitar. The colums "Separated instruments", "Separated playing techniques", and "Separated pick-up setting" indicate which criteria were applied to separate the samples from training and test set in each configuration. The fifth and sixth columns indicate whether the plausibility filter (compare Section 4.3) and the frame result aggregation (compare Section 4.3) were applied. In the seventh column, the number of folds for the configuration 1.6 and 2.6 and the number of permutations for the remaining configurations are given. The evaluation measures precision, recall, and F-measure were always averaged over all permutations and all folds, respectively.

After the training set and the test set were separated, the columns of the training feature matrix were first normalized to zero mean and unit variance. The mean vector and the variance vector were kept for later normalization of the test data. Subsequently, the normalized training feature matrix was used to derive the transformation matrix via LDA. The SVM model was then trained using the projected training feature matrix and a two-dimensional grid search is performed to determine the optimal parameters C and γ as explained in 4.3. For the configurations 1.6 and 2.6, none of the criteria to separate the training

Table 3. Mean Precision $\bar{P}$, mean Recall $\bar{R}$, and mean F-Measure $\bar{F}$ for different evaluation conditions (compare Section 5.2) for the bass guitar

Experiment	Separated instruments	Separated playing techniques	Separated pick-up settings	Plausibility filter (see Section 4.3)	Result aggregation over 5 frames (see Section 4.3)	No. of Permutations$^{\diamond}$/ No. of CV folds*	Precision $\bar{P}$	Recall $\bar{R}$	F-Measure $\bar{F}$
1.1.a	x					$2^{\diamond}$	.85	.85	.85
1.1.b	x			x		$2^{\diamond}$	.87	.87	.87
1.1.c	x			x	x	$2^{\diamond}$	.78	.78	.78
1.2.a	x	x				$8^{\diamond}$	.86	.86	.86
1.2.b	x	x		x		$8^{\diamond}$	.87	.87	.87
1.2.c	x	x		x	x	$8^{\diamond}$	.88	.88	.88
1.3.a		x	x			$8^{\diamond}$	.57	.50	.49
1.3.b		x	x	x		$8^{\diamond}$	.71	.69	.69
1.3.c		x	x	x	x	$8^{\diamond}$	.88	.88	.88
1.4.a		x				$8^{\diamond}$	.60	.54	.54
1.4.b		x		x		$8^{\diamond}$	.73	.71	.72
1.4.c		x		x	x	$8^{\diamond}$	**.93**	**.93**	**.93**
1.5.a			x			$8^{\diamond}$	.62	.55	.54
1.5.b			x	x		$8^{\diamond}$	.74	.71	.71
1.5.c			x	x	x	$8^{\diamond}$	.92	.92	.92
1.6.a						$10^{\star}$	.92	.92	.92
1.6.b				x		$10^{\star}$	.93	.93	.93
1.6.c				x	x	$10^{\star}$	**.93**	**.93**	**.93**

and the test set was applied. Instead, here we used a 10-fold cross-validation and averaged the precision, recall, and F-measure over all folds.

The results shown in Table 3 and Table 4 clearly show that both the plausibility filter as well as the result aggregation step significantly improved the classification results in most of the investigated configurations. Furthermore, it can be seen that the separation of training and test samples according to instrument, playing technique, and pick-up setting has a strong influence on the achievable classification performance. In general, the results obtained for the bass guitar and the electric guitar show the same trends. We obtain the highest classification scores—$\bar{F} = .93$ for the bass guitar (4 classes) and $\bar{F} = .90$ for the electric guitar (6 classes)—for the configurations 1.6 and 2.6. These results indicate that the presented method can be successfully applied in different application tasks that require an automatic estimation of the played instrument string. In contrast to [17], we did not make use of any knowledge about the musical context such as that which may be derived from a musical score.

Experiment 3: Baseline Experiment Using MFCC Featues. We performed a baseline experiment separately for both instruments using Mel Fre-

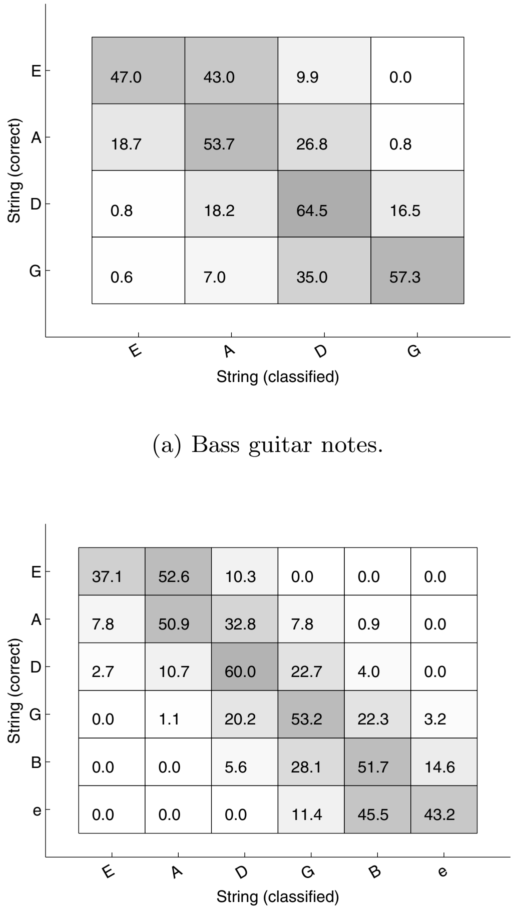
Table 4. Mean Precision $\bar{P}$, mean Recall $\bar{R}$, and mean F-Measure $\bar{F}$ for different evaluation conditions (compare Section 5.2) for the electric guitar

Experiment	Separated instruments	Separated playing techniques	Separated pick-up settings	Plausibility filter (see Section 4.3)	Result aggregation over 5 frames (see Section 4.3)	No. of Permutations$^\diamond$/ No. of CV folds*	Precision $\bar{P}$	Recall $\bar{R}$	F-Measure $\bar{F}$
2.1.a	x					$2^\diamond$	.64	.64	.63
2.1.b	x			x		$2^\diamond$	.70	.70	.70
2.1.c	x			x	x	$2^\diamond$	.76	.75	.75
2.2.a	x	x				$8^\diamond$	.69	.69	.68
2.2.b	x	x		x		$8^\diamond$	.71	.71	.70
2.2.c	x	x		x	x	$8^\diamond$	.78	.77	.77
2.3.a		x	x			$8^\diamond$	.61	.57	.56
2.3.b		x	x	x		$8^\diamond$	.68	.66	.66
2.3.c		x	x	x	x	$8^\diamond$	.74	.74	.73
2.4.a		x				$8^\diamond$	.64	.61	.60
2.4.b		x		x		$8^\diamond$	.71	.69	.69
2.4.c		x		x	x	$8^\diamond$	.80	.79	.79
2.5.a			x			$8^\diamond$	.69	.65	.65
2.5.b			x	x		$8^\diamond$	.74	.72	.72
2.5.c			x	x	x	$8^\diamond$	.84	.84	.84
2.6.a						$10^\star$	.72	.69	.70
2.6.b				x		$10^\star$	.81	.81	.81
2.6.c				x	x	$10^\star$	**.90**	**.90**	**.90**

quency Cepstral Coefficients (MFCC) as features. Again, LDA and SVM were applied for feature space transformation and classification respectively, as explained in Section 4.3. The same experimental conditions as in configuration 1.6. and 2.6. (see 5.2) were used. The classification results were performed and evaluated on a frame level. A 10-fold stratified cross-validation was applied and the results were averaged over all folds. We achieved classification scores of $\bar{F} = .46$ for the bass guitar and $\bar{F} = .37$ for the electric guitar.

Experiment 4: Human Performance on String Classification. In the final experiment, we aim to investigate the performance of human listeners for the given task of classifying the string number based on isolated bass guitar and electric guitar notes. The study comprised 9 participants, most of them being semi-professional guitar or bass players. To allow for a comparison between the algorithm performance and the human performance, similar test conditions must be guaranteed. Based on the results shown in Table 3 and Table 4, the conditions of Experiments 1.6.c and 2.6.c are used for the listening test. The samples are randomly assigned to training and test set—no separation based on playing technique, pick-up setting, or instrument is performed.

During the training phase, the participants could listen to as many notes from the training set for each string class as they wanted to. Afterwards, they were asked to assign randomly selected samples from the test set to one of the 4 or 6 string classes, respectively. Overall, 578 guitar notes and 522 bass notes were annotated with a string number.



String (correct) \ String (classified)	E	A	D	G
E	47.0	43.0	9.9	0.0
A	18.7	53.7	26.8	0.8
D	0.8	18.2	64.5	16.5
G	0.6	7.0	35.0	57.3

(a) Bass guitar notes.

String (correct) \ String (classified)	E	A	D	G	B	e
E	37.1	52.6	10.3	0.0	0.0	0.0
A	7.8	50.9	32.8	7.8	0.9	0.0
D	2.7	10.7	60.0	22.7	4.0	0.0
G	0.0	1.1	20.2	53.2	22.3	3.2
B	0.0	0.0	5.6	28.1	51.7	14.6
e	0.0	0.0	0.0	11.4	45.5	43.2

(b) Electric guitar notes.

Fig. 5. Confusion matrix from human performance for string classification. All values are given in percent.

As it can be seen in the two confusion matrices in Figure 5, human listeners tend to confuse notes between adjacent strings on the instrument. In total, classification scores of $\bar{F} = .27$ for the bass guitar and $\bar{F} = .16$ for the electric guitar were achieved.

6 Conclusions

In this paper, we performed several experiments geared towards the automatic classification of the string number from given isolated note recordings. We presented a selection of audio features that can be extracted on a frame-level. In order to improve the classification results, we first applied a plausibility filter to avoid non-meaningful classification results. Then, we used an aggregation of multiple classification results obtained from adjacent frames of the same note. Highest string classification scores of $\bar{F} = .93$ for the bass guitar (4 string classes) and $\bar{F} = .90$ for the electric guitar (6 string classes) were achieved. As shown in a baseline experiment, classification systems based on commonly-used audio features such as MFCC were clearly outperformed for the given task. The task of automatic string detection is very challenging for human listeners as the results of a listening tests confirmed: F-measure values of only $\bar{F} = .27$ and $\bar{F} = .16$ could be achieved for the bass guitar and the electric guitar, respectively.

7 Outlook

7.1 String Detection for Melodies and Chords

As mentioned in Section 2, guitar players and bass players usually choose the fingering to play a given music score in such a way that the overall physical strain is minimized. One major characteristic of this behavior is the preference of vertical play over horizontal play on the instrument neck: instead of playing melodies only on one or two adjacent strings with a strong vertical hand movement over the instrument neck, musicians prefer to stay in a fixed fretboard position as long as possible and try to use the whole possible pitch range, available there. This knowledge could be used to implement a temporal modeling of fretboard position changes over time using a Hidden-Markov Model (HMM) or a comparable method.

Secondly, polyphonic music signals such as chords played on a guitar were not covered in this paper. Multi-pitch estimation is still one of the most challenging tasks in Music Information Retrieval [7,15]. In order to apply the feature-based approach for string detection as presented in this paper, several challenges need to be overcome. Guitar chords often contain up to 6 simultaneous sounding notes and furthermore include many octave intervals. As a consequence, many of the harmonics overlap in the frequency domain. This impedes the precise estimation of the harmonic frequency values and the computation of meaningful audio features (compare 4.2). Fortunately, all notes in guitar chords are always played within a fixed fret range on the instrument neck due to the limited span of the human hand. Thus, even if the string classification results are erroneous for some notes of a chord, the use of a majority voting scheme could be applied over all single string classification results to get a robust estimate of the overall fretboard position of an analyzed chord.

7.2 Application of String Classification for Music Education Software

In the context of music education software, the main application of string classification is to automatically evaluate how well a musician follows a given tablature. String classification can therefore be interpreted as an extension to conventional music transcription systems that is tailored for the analysis of string instrument performances. Assuming that the system can initially be trained with a dataset comparable to the one used in this paper, the system performance on notes from a different instrument[6] will likely be as in Experiment 1.1. or 2.1. since in these experiments, different instruments are used for training and testing.

In order to improve classification results, the *online learning* paradigm can be applied here: by using the software with his or her instrument, the user will continuously provide new training data to the system. The program can adapt to the timbre of the applied instrument by taking a selection of the recorded instrument notes. The class label of each new note can be taken from the corresponding playing instruction in the program. After adapting to the new instrument, the classification results will likely achieve values comparable to Experiment 1.6. or 2.6. In the context of a music learning application, no strict separation between training and test data is necessary, and “overfitting” to the player’s instrument is reasonable and beneficial for the system’s overall performance. However, occasional playing errors that involve playing on the wrong string could lead to the selection of training samples with erroneous string class labels. Those training samples could corrupt the improvements gained from using online learning.

Another challenge that was not covered in this paper is that in practice, strings with different material and gauges are used for different music styles. The string gauge directly affects the string inharmonicity coefficient and thus the features used for string classification. If the training samples for each string class are recorded with instruments having too many different string gauges, the class distributions in the feature space will overlap more and the classification performance is expected to decrease.

Acknowledgements. The author author would like to thank all participants of the listening tests as well as Michael Stein for the use of his data set. The Thuringian Ministry of Economy, Employment and Technology supported this research by granting funds of the European Fund for Regional Development to the project *Songs2See*[7], enabling transnational cooperation between Thuringian companies and their partners from other European regions.

[6] Since the musician uses his or her own instrument, it is most likely that this instrument is not incorporated in the initial training set.

[7] `http://www.songs2see.net`

References

1. Abeßer, J., Dittmar, C., Schuller, G.: Automatic Recognition and Parametrization of Frequency Modulation Techniques in Bass Guitar Recordings. In: Proc. of the 42nd Audio Engineering Sociery (AES) International Conference on Semantic Audio, Ilmenau, Germany, pp. 1–8 (2011)
2. Barbancho, A.M., Klapuri, A., Tardón, L.J., Barbancho, I.: Automatic Transcription of Guitar Chords and Fingering from Audio. IEEE Transactions on Audio, Speech, and Language Processing, 1–19 (2011)
3. Barbancho, I., Barbancho, A.M., Tardón, L.J., Sammartino, S.: Pitch and Played String Estimation in Classic and Acoustic Guitars. In: Proceedings of the 126th Audio Engineering Society (AES) Convention, Munich, Germany (2009)
4. Barbancho, I., Member, S., Tardón, L.J., Sammartino, S., Barbancho, A.M.: Inharmonicity-Based Method for the Automatic Generation of Guitar Tablature. IEEE Transactions on Audio, Speech, and Language Processing 20(6), 1857–1868 (2012)
5. Burns, A., Wanderley, M.: Visual Methods for the Retrieval of Guitarist Fingering. In: Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME 2006), Paris, France, pp. 196–199 (2006), `http://portal.acm.org/citation.cfm?id=1142263`
6. Chang, C.C., Lin, C.J.: LIBSVM: A Library for Support Vector Machines. Tech. rep., Department of Computer Science, National Taiwan University, Taipei, Taiwan (2011)
7. Christensen, M.G., Jakobsson, A.: Multi-Pitch Estimation. Synthesis Lectures on Speech and Audio Processing, Morgan & Claypool Publishers (2009)
8. Fletcher, N.H., Rossing, T.D.: The Physics of Musical Instruments, 2nd edn. Springer, New York (1998)
9. Galembo, A., Askenfelt, A.: Measuring inharmonicity through pitch extraction. Speech Transmission Laboratory. Quarterly Progress and Status Reports (STL-QPSR) 35(1), 135–144 (1994)
10. Galembo, A., Askenfelt, A.: Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano. IEEE Transactions on Speech and Audio Processing 7(2), 197–203 (1999)
11. Hodgkinson, M., Timoney, J., Lazzarini, V.: A Model of Partial Tracks for Tension-Modulated Steel-String Guitar Tones. In: Proc. of the 13th Int. Conference on Digital Audio Effects (DAFX 2010), Graz, Austria, pp. 1–8 (2010)
12. Hrybyk, A., Kim, Y.: Combined Audio and Video for Guitar Chord Identification. In: Proc. of the 11th International Society for Music Information Retrieval Conference (ISMIR), Utrecht, Netherlands, pp. 159–164 (2010)
13. Järveläinen, H., Välimäki, V., Karjalainen, M.: Audibility of the timbral effects of inharmonicity in stringed instrument tones. Acoustics Research Letters Online 2(3), 79 (2001)
14. Kerdvibulvech, C., Saito, H.: Vision-Based Guitarist Fingering Tracking Using a Bayesian Classifier and Particle Filters. In: Mery, D., Rueda, L. (eds.) PSIVT 2007. LNCS, vol. 4872, pp. 625–638. Springer, Heidelberg (2007)
15. Klapuri, A.: Multipitch analysis of polyphonic music and speech signals using an auditory model. IEEE Transactions on Audio, Speech, and Language Processing 16(2), 255–266 (2008)
16. Lukashevich, H.: Feature selection vs. feature space transformation in automatic music genre classification tasks. In: Proc. of the AES Convention (2009)

17. Maezawa, A., Itoyama, K., Takahashi, T., Ogata, T., Okuno, H.G.: Bowed String Sequence Estimation of a Violin Based on Adaptive Audio Signal Classification and Context-Dependent Error Correction. In: Proc. of the 11th IEEE International Symposium on Multimedia (ISM 2009), pp. 9–16 (2009)
18. Marple, S.L.: Digital Spectral Analysis With Applications. Prentice Hall, Australia (1987)
19. O'Grady, P.D., Rickard, S.T.: Automatic Hexaphonic Guitar Transcription Using Non-Negative Constraints. In: Proc. of the IET Irish Signals and Systems Conference (ISSC), Dublin, Ireland, pp. 1–6 (2009)
20. Paleari, M., Huet, B., Schutz, A., Slock, D.: A Multimodal Approach to Music Transcription. In: Proc. of the 15th IEEE International Conference on Image Processing (ICIP), pp. 93–96 (2008)
21. Peeters, G., Rodet, X.: Hierarchical gaussian tree with inertia ratio maximization for the classification of large musical instruments databases. In: Proc. of the Int. Conf. on Digital Audio Effects (DAFx), London, UK (2003)
22. Penttinen, H., Siiskonen, J.: Acoustic Guitar Plucking Point Estimation in Real Time. In: Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 209–212 (2005)
23. Saito, S., Kameoka, H., Takahashi, K., Nishimoto, T., Sagayama, S.: Specmurt Analysis of Polyphonic Music Signals. IEEE Transactions on Audio, Speech, and Language Processing 16(3), 639–650 (2008)
24. Stein, M., Abeßer, J., Dittmar, C., Schuller, G.: Automatic Detection of Audio Effects in Guitar and Bass Recordings. In: Proceedings of the 128th Audio Engineering Society (AES) Convention, London, UK (2000)
25. Välimäki, V., Pakarinen, J., Erkut, C., Karjalainen, M.: Discrete-time modelling of musical instruments. Reports on Progress in Physics 69(1), 1–78 (2006)
26. Vapnik, V.N.: Statistical learning theory. Wiley, New York (1998)